

Assignment 1

Ivan Aristy — iae225

October 2, 2024

1 Assignment

1.1 Designing a Static Visualization

Your task for this assignment is to design a static visualization that you believe effectively communicates an idea or message about the sunshine data, and provide a written report (no more than 1 page, single spaced, not including images) that details your design. Note that you should include at least one image of your visualization in your report (i.e., when all of the images are excluded from the report, the text does not exceed 1 page). We recommend that you start this assignment by identifying a question about the data that you would like your visualization to answer. Then, you should design a visualization to answer that question, and use that question as the title of your visualization.

You are required to use the given data set, but you are free to transform the data in any way that you wish. That is, you can manipulate the data using transformations such as a log transformation, computing percentages or averages, grouping elements into categories, or removing unnecessary variables. You are also allowed to incorporate additional data from external sources, but if you do so, you need to make a note of it in your submission.

Remember that different visualizations can emphasize different aspects of a data set, so your writeup should include details about which components of the data you intended to communicate. You should also provide comments on which aspects of the data set are obscured by your visualization design.

Your report should explain the rationale behind your design decisions. You should document the visual encodings you used and why they are effective for communicating your intended message. These decisions include things like the choice of visualization type, size, color, scale, and other visual elements (which we will learn more about during Week #3 and Week #4 lectures).

Note, the text for this document is under a standard word page when excluding the images and pretty formatting.

1.2 Intended Visualization & Question

During which months are cities between the 68 and 80°F range, while maximizing for sunshine hours?

This is the question we aim to answer in this assignment. The reason for this question is that the 68-80°F range is considered the most comfortable for most people, and sunshine during this time will make living there a better experience. The audience would be looking for the best cities climate-wise for a vacation, to move to, or maybe a second house.

1.3 Data Cleaning & Gathering

We first need to consider if we have all the elements necessary to answer this question. In particular, we already have sunshine hours per month, and the particular cities we want to analyze. However, we do not have the temperature data for each city. Hence, we can retrieve this data from weather.gov

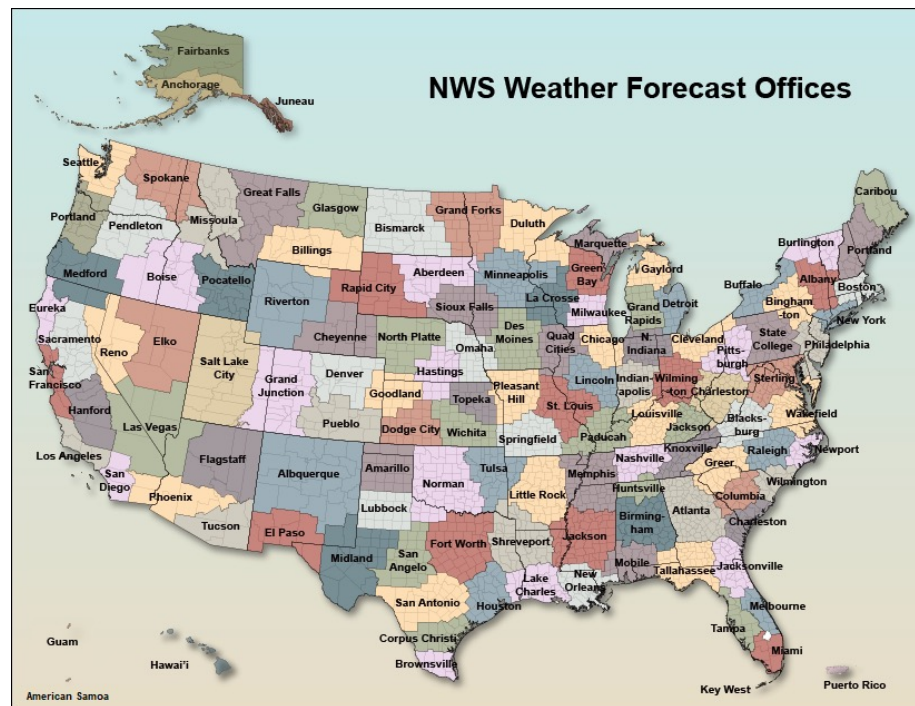


Figure 1: Map of the United States showing available data.

NOWData - NOAA Online Weather Data

1. Location »

[View map](#)

- Miami Area
- West Palm Beach Area
- Fort Lauderdale Area
- Brighton, FL
- Canal Point Usda, FL
- Cape Florida, FL
- Devils Garden, FL
- Flamingo Ranger, FL
- Fort Lauderdale, FL
- Ft Lauderdale Bch, FL

2. Product »

- ☐ Daily data for a month
- ☐ Daily almanac
- ☒ Monthly summarized data
- ☐ Calendar day summaries
- ☐ Daily/monthly normals
- ☐ Climatology for a day
- ☐ First/last dates
- ☐ Temperature graphs
- ☐ Accumulation graphs

3. Options »

Year range: 2023 - 2023

Variable

Avg temp

Summary

Mean


4. View »

[Go](#)

Product Description:

MONTHLY SUMMARIZED DATA - calculates averages, totals, daily extremes or frequencies for the selected variable for each month of the year for the selected range of years. Note: trace precipitation/snowfall/snow depth amounts are treated as zero in sums, mean, and number of days counts. Annual average temperatures are the average of the twelve monthly values. Temperatures are reported in degrees F; precipitation, snowfall and snow depth are reported in inches.

[- Common questions -](#)



The Applied Climate Information System (ACIS) is a joint project of the Regional Climate Centers, the National Centers for Environmental Information (NCEI) and the National Weather Service. Official data and data for additional locations are available from the Regional Climate Centers and NCEI.

Figure 2: Data request for weather.gov

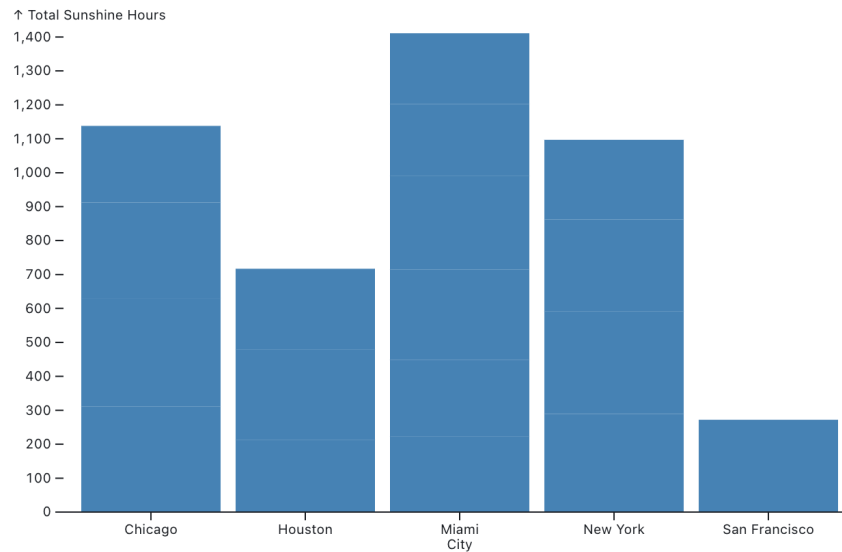
NOWData - NOAA Online Weather Data														Enlarge results	Print	X
Monthly Mean Avg Temperature for Miami Area, FL (ThreadEx)																
Click column heading to sort ascending, click again to sort descending.																
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual			
2023	72.1	75.9	76.8	79.4	81.2	83.7	86.5	86.6	84.7	82.0	77.8	71.2	79.8			
Mean	72.1	75.9	76.8	79.4	81.2	83.7	86.5	86.6	84.7	82.0	77.8	71.2	79.8			
Max	72.1	75.9	76.8	79.4	81.2	83.7	86.5	86.6	84.7	82.0	77.8	71.2	79.8			
Min	72.1	75.9	76.8	79.4	81.2	83.7	86.5	86.6	84.7	82.0	77.8	71.2	79.8			

Figure 3: Obtained data for temperature.

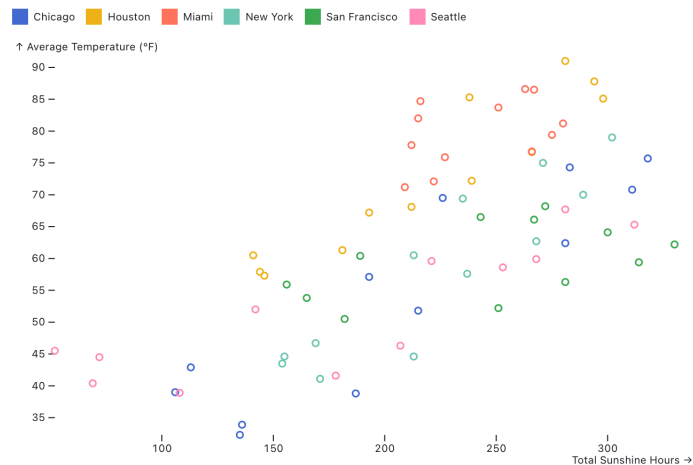
Also, we do not need to consider latitude and longitude for our visualization, so we can remove or ignore these columns.

1.4 Iterations

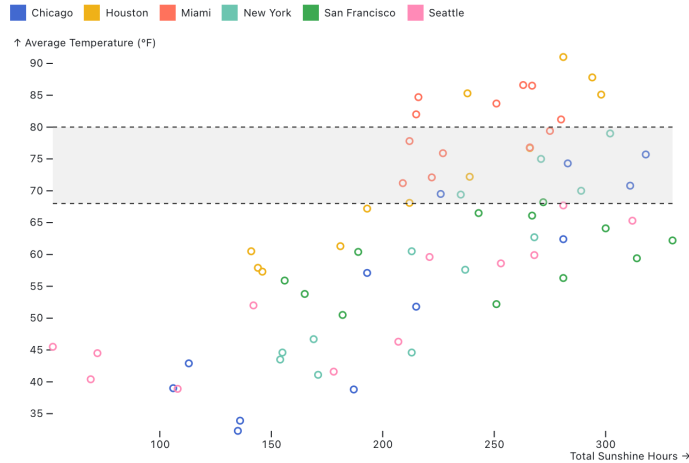
First I tried to filter the data to only show cities with a temperature between 68 and 80°F:



Ignore, that there's no visual encoding, this is just testing. However, I think this approach is not effective in communicating the data. We could instead try to encode the data using a scatter plot, with temperature on the x-axis, and sunshine hours on the y-axis.



Adding a highlighted section for the 68-80°F range:



This shows more information, but we have (a) non-clarification of months, which is crucial for the audience, and (b) it feels cluttered.

The major issue is that we are trying to represent 4 variables in a 2D space. I also considered a bubble chart, where the x-axis represents months, the y-axis represents the cities, the size of each bubble represents sunshine hours, and the color of the bubbles represents the temperature. This would no doubt feel cluttered, and the audience would have a hard time interpreting the data.

Hence, the two best approaches are to either filter the data to only show cities in the 68-80°F range in a line chart, or to add two panels for the dependent vars:

1.5 Final Visualization

1.5.1 Does this answer the question?

We highlight some key data:

- **Sunshine** lorem ipsum dolor sit amet
- **Temperature** lorem ipsum dolor sit amet
- **City** lorem ipsum dolor sit amet

1.5.2 Encoding (Marks/Channels, Redundancy, Color/Accessibility)

1.5.3 Obscured Data

- **Preference:** The visualization is not dynamic, hence we cannot allow the user to select their desired temperature range.

- **Lat-Long:** Our visualization assumes that the audience is familiar with the cities in the dataset, and doesn't provide a map for reference.
- **Humidity:** We do not consider humidity in our visualization, which could be a factor in the comfort of the weather.